91602
new 82-10 LEGAL NOTICE

This report was prepared by Ontario Hydro and, while effort has been made to assure the accuracy and applicability of the material contained herein, neither Ontario Hydro nor any person acting on its behalf:

(a) makes any warranty, express or implied with respect to the use of any information, apparatus, method or process disclosed in this report or that such use may not infringe privately owned rights; or

(b) assumes any liabilities with respect to the use of, or for damages resulting from the use of, any information, apparatus, method or process disclosed in this report.

To: Dr. D. Mills
Research Division

ontario hydro
research division

COMPLIMENTARY COPY
Please Credit Ontario Hydro Research if Reproduced in Whole or in Part

ZERO COUNTS IN FISH ENTRAINMENT SURVEYS

S.A. Bartlett
Mathematician
Reliability & Statistics Section
Operations Re[illegible] Department

ABSTRACT

Zero counts in fish entrainment surveys occur frequently but can lead to large variation in the estimates of total entrainment. One objective in designing a survey is, therefore, to minimize the probability of obtaining zero counts in the data. A first step towards meeting this objective is to determine the causes of zero counts. A small study was carried out at Pickering NGS in 1988 to help in this respect. Analysis showed that zero counts can occur due to short sample times, small numbers of larvae or eggs, and that larvae or eggs occur in patches. The Poisson distribution and the delta-lognormal mixing distributions appear to describe the distribution of larvae or eggs very well. Increasing the sampling times contributes to reducing the probability of obtaining zero counts.

1. INTRODUCTION

Fish entrainment surveys are conducted at power plant intakes in order to assess the impact of such entrainment on fish stocks in the area. The usual method is to sample fish larvae and eggs flowing into the intake periodically during the larvae and egg seasons, anytime from early spring to fall. A conical net is dropped in the water and the larvae or eggs are caught as the water flows through. The nets are kept in the water for up to thirty minutes. Often the samples do not contain any larvae or eggs of the species of fish of interest, which leads to zero counts. Zero counts can lead to large variation in the estimates of total entrainment obtained from the survey.

job	file	date	report no.
72-510-001	809.7	July 18, 1989	89-178-K

The Biology Section of the Research Division conducts many surveys for estimating fish abundances and would like to know how to minimize this variation. In the summer of 1988, they performed a small survey to collect data to study this further. The Statistics Unit was requested to perform an analysis of the data. This report contains the results of that analysis and recommendations.

In a study of entrainment sampling procedures done by the Statistics Unit for the Environmental Studies and Assessments Department/1/, it was determined that possible explanations for zero counts are that the sample times are not long enough to detect larvae or eggs, that the numbers are so small that zero counts were probable, or that the larvae and eggs occur in patches, some of which are more dense than others. It is suspected that the true explanation is a combination of all three possibilities.

The delta-Poisson-lognormal distribution/2/ describes this behavior quite well. The Poisson part comes from the fact that count data per unit effort such as larvae catch in a ten minute tow usually have a Poisson distribution. The lognormal part takes into consideration the large spatial and temporal variation. The delta part helps to account for zero counts that arise from the movements of the patches. Note that this does not include the zeros that could arise from the Poisson distribution.

Two questions arise:

1. What zero counts are due to the Poisson part of the distribution and what part are due to the delta part of the distribution?

2. What impact does the sampling time have on the probability of getting zeros?

2. DATA

The study was conducted at Pickering Nuclear Generation Station in the summer of 1988. Sampling of fish larvae was carried out using 0.5 meter (m) conical nets at two points on a deck approximately 1 to 1-1/2 meters apart. Ten minute samples of the water flowing into the intake were taken on four consecutive days, July 27 through July 30. On July 27, five samples were taken starting at 14:54 hours with one to two minute intervals in between samples. On the remaining three days, three sequences of five samples were taken at approximately 09:00, 12:00, and 15:00 hours. Sampling began when the biologists noticed peak larval densities occurring. The number of larvae in each sample were then counted. No attempt was made to distinguish between live and dead larvae. The water flowing into the intake was considered to be constant at 96.56 m^3/10 min throughout the study.

Table 1 contains a list of the collected data. There was only one count of zero for tow 1, July 30 at 09:23 hours. There were problems associated with some of the samples: the samples were incorrectly observed, the bucket tipped, or the count was only of dead larvae since live larvae were not properly preserved. These samples were eliminated from the analysis. In one case, the sample was redone at a later time (tow 2 on July 28 at 15:33 hours). The two samples taken at 12:12 hours on July 30 were from twenty minute tows rather than ten minute tows. So that the counts conform to the rest of the data, the numbers were divided by two to yield new counts of 1.5 larvae for tow 1 and 0.5 larvae for tow 2. Similarly, the two samples taken at 12:46 hours on July 30 were from twelve minute tows. The counts were thus multiplied by 10/12 to yield new counts of 4.2 larvae for tow 1 and 5.0 larvae for tow 2.

Figure 1 contains a plot of counts in each sample versus the replicates in the order they were gathered. Sequences 1, 2, and 3 refer to the 09:00, 12:00 and 15:00 hour sequences, respectively. Generally, the replicate counts within sequence and tow combinations appear to have less variation relative to counts between sequence and tow combinations. There does not appear to be much difference between tows. Some sequence and tow combinations have large variation such as sequence 1, tow 2 on July 29 where the counts are decreasing rapidly.

The variation in counts appears to decrease for smaller counts. This is expected since count data usually has a Poisson distribution and the variance depends on the magnitude of the mean. The larger the number, the larger the variance.

3. ANALYSIS

3.1 Dependence on Sampling Times

To determine the impact of the sampling time on the probability of getting zero counts in a sample, it will be assumed that the counts have strictly a Poisson distribution. That is, if x is the number of larvae in the sample, then the probability that x equals a particular number (n) is

$$\text{Prob}\ (x=n) = e^{-\mu}\mu^n/n!$$

where μ is the *expected* number of larvae in the sample/3/. Note that the *expected* number is different from the *observed* number (x) in that it is a constant, theoretical value whereas the *observed* number has variation associated with it.

The probability of obtaining a count of zero given that the *expected* counts are specified is calculated for particular sampling times, then recalculated by increasing or decreasing the sampling times. The results will give an idea of the impact the sampling time has on detecting larvae.

In Table 2, the probabilities are given for various numbers of larvae *expected* to be present in a ten-minute sample. The *expected* counts range from one to ten. Twenty is given for an indication of what happens for larger *expected* counts. If the sample time is actually ten minutes, then the probability of finding no larvae in the sample is 0.37 given that the *expected* number is one every ten minutes (see the column under ten minutes sample time). If the *expected* number is two in ten minutes, then the probability of finding no larvae in the sample decreases to 0.13. As the *expected* count increases, the chance of finding no larvae in a 10 minute sample drops off rapidly.

If the *expected* number of larvae in a ten-minute sample is one, but the sample was only taken for five minutes, then the expected number in the five-minute sample becomes 0.5 and the probability of finding no larvae in the tow jumps to 0.61 (see the column of values under five-minute samples). If the sample was only ten for one minute, the probability of finding no larvae increases to 0.91. On the other hand, increasing the sampling times decreases the probability of getting zero counts (see columns under sample times of fifteen and twenty minutes).

From the above discussion it is clear that the sampling time affects the probability of getting zero counts in the sample given that the counts have a Poisson distribution. To reduce the probability of getting a zero count, the sampling time should be increased. If the researcher knows generally what kind of counts can be expected in the high density part of the season, then it is relatively easy to chose an appropriate sampling time.

3.2 Separation of Poisson and Lognormal Distributions

The first question in Section 1 cannot be directly addressed since the data only had one zero count, an insufficient number for rigorous analysis. However, it is also of interest to be able to determine that portion of the variation which can be attributed to the Poisson distribution and the portion attributed to the lognormal distribution. It is beyond the scope of this study to do a rigorous analysis since extensive development of the model is required. Rather, some remarks will be made to give direction to future study.

Table 3 contains the means and variances of the replicates within each sequence and tow combination. If the counts have a Poisson distribution, then the mean and variance should be equivalent/3/. The data for the most part appear to be consistent with the Poisson distribution since the means and variances are roughly of the same magnitude. A few sequence and tow combinations have variances much larger than the means such as in sequence 1, tow 1 on July 28, where the mean of the five replicates is 44.2 while the variance is 301.7. This may mean that the lognormal mixing distribution is playing a larger role in some sequence and tow combinations than in others.

As mentioned earlier, there was only one sample with a zero count. It can be commented on whether or not this is consistent with the idea of the counts having a Poisson distribution within the sequence. To do this, the probability of obtaining at least one zero in the five replicates is determined. First, the probability is computed that a particular sample is zero given that the expected mean is equal to the mean (m) of the replicates within the sequence and tow combination. This is equivalent to

$$p = \text{Prob}(x=0) = e^{-m}m^0/0! = e^{-m}$$

Second, this value is used to calculate the probability that at least one replicate out of five will be zero using binomial probabilities. Thus, the probability is

$$\text{Prob(at least one zero)} = 1 - \binom{5}{0} p^0(1-p)^5$$

$$= 1 - (1-p)^5$$

These values are contained in Table 3 for all sequence and tow combinations. Assuming the counts have a Poisson distribution, the probability of getting at least one zero in the first sequence of the first tow on July 30 is 0.2254. This is relatively high so that getting a zero count is not unexpected. Other combinations had probabilities of getting zero counts greater than 0.05 (all on July 30). Yet no samples had zero counts. The probability of getting a zero count is less than 0.3.

It appears that the variation within a sequence and tow combination can be described for the most part by the Poisson distribution. However, there appears to be more variation between sequences that is probably due to other sources of variation. Analysis of variance was used to determine if there were significant differences between tows, among sequences and among days.

The analysis was performed on the square root of the counts. This was done because the variance of the counts increases with magnitude of the counts thereby violating a fundamental criteria for using the analysis of variance procedure/4/, namely, that the variance is uniform for all data points. In addition, the third sequence of July 27 was omitted since it was the only sequence on that day and including it would unbalance the analysis, thereby producing less statistically reliable results.

The results of the analysis of variance on the square roots of the counts is shown in Table 4. Note that the interaction between day and sequence is included as a factor. Other interactions were omitted since there was no physical reason to suspect they existed. Tow is not significant though day, sequence, and their interaction are. The variance due to the factors is 35.64 while the within sequence variance is 1.03 (see mean square errors for the model and error sources in Table 4, respectively). Thus, differences among sequences and days describe most of the variation in the count data while within sequence describes relatively little of the variation. If tow is omitted, similar conclusions are reached (Table 5).

Diagnostics on the residuals, the differences between the observed counts and counts predicted by the analysis of variance model, were done to ensure that assumptions on which the reliability of analysis depends are not violated. These assumptions are that the residuals are independently, normally distributed with uniform variance. There was no evidence to indicate that the residuals were not independently and normally distributed. Some of the the variances of the sequence and tow combinations were quite large compared to others. However, analysis of variance works quite well even if the ratio of the largest variance to the smallest of the sequence and tow combinations is as high as twenty/5/. With the exception of one variance, the ratio of the largest to smallest falls in this range.

There are bound to be differences in mean counts among days since counts are dependent on the day of the season. However, differences among sequences within days may not be as expected. Separate analyses were done for each day excluding July 27 and omitting tow. Sequence was significant for July 28 and 29, but not for July 30 (see Table 6). There is great variation throughout the day indicating the spatial and temporal variation is coming into play. Thus, it appears that the lognormal distribution needs to be used to describe the variation throughout the day.

It is noted that of possible auto correlation between replicate counts within sequence and tow combinations is ignored. It is impossible to include this effect in the analysis with this data since the sequences are not long enough.

4. CONCLUSIONS

- Sampling times have a large effect on the probability of getting a zero count in the sample.
- The two tows are not significantly different from each other.
- For samples taken at times close together, the counts are more uniform than those taken at different times during the day.
- The variation of samples taken close together is likely described by the Poisson distribution.
- The temporal and spatial variation of samples taken at different times of the day is likely described by the delta-lognormal distribution.

5. RECOMMENDATIONS

- To avoid zero counts, sample times should be long enough that the probability of getting zero counts is minimized. This may not be possible for species with low incidence such as whitefish.

Approved:

A.H. Chung
Manager
Operations Research Department

Submitted:

S.A. Bartlett
Mathematician
Statistics & Reliability Section

SAB:cmh

REFERENCES

1. Bartlett, S.A., "Sampling Schemes for Fish Entrainment Studies", Ontario Hydro Research Division report (draft).

2. Crow, E.L., and Shimizu, K. (eds.), *Lognormal Distributions Theory and Applications*, Marcel Dekker, New York, 1988, p. 324-326.

3. McClave, J.T., and Dietrich, F.H., *Statistics*, Dellan Publishing Co., San Francisco, Calif, 1979, p. 143.

4. Rao, C.R., *Linear Statistical Inference and Its Applications*, 2nd Ed, John Wiley & Sons, New York, 1973, p. 426.

5. Green, R.H., *Sampling Design and Statistical Methods for Environmental Biologists*, John Wiley & Sons, New York, 1979, p.44.

FIGURE 1

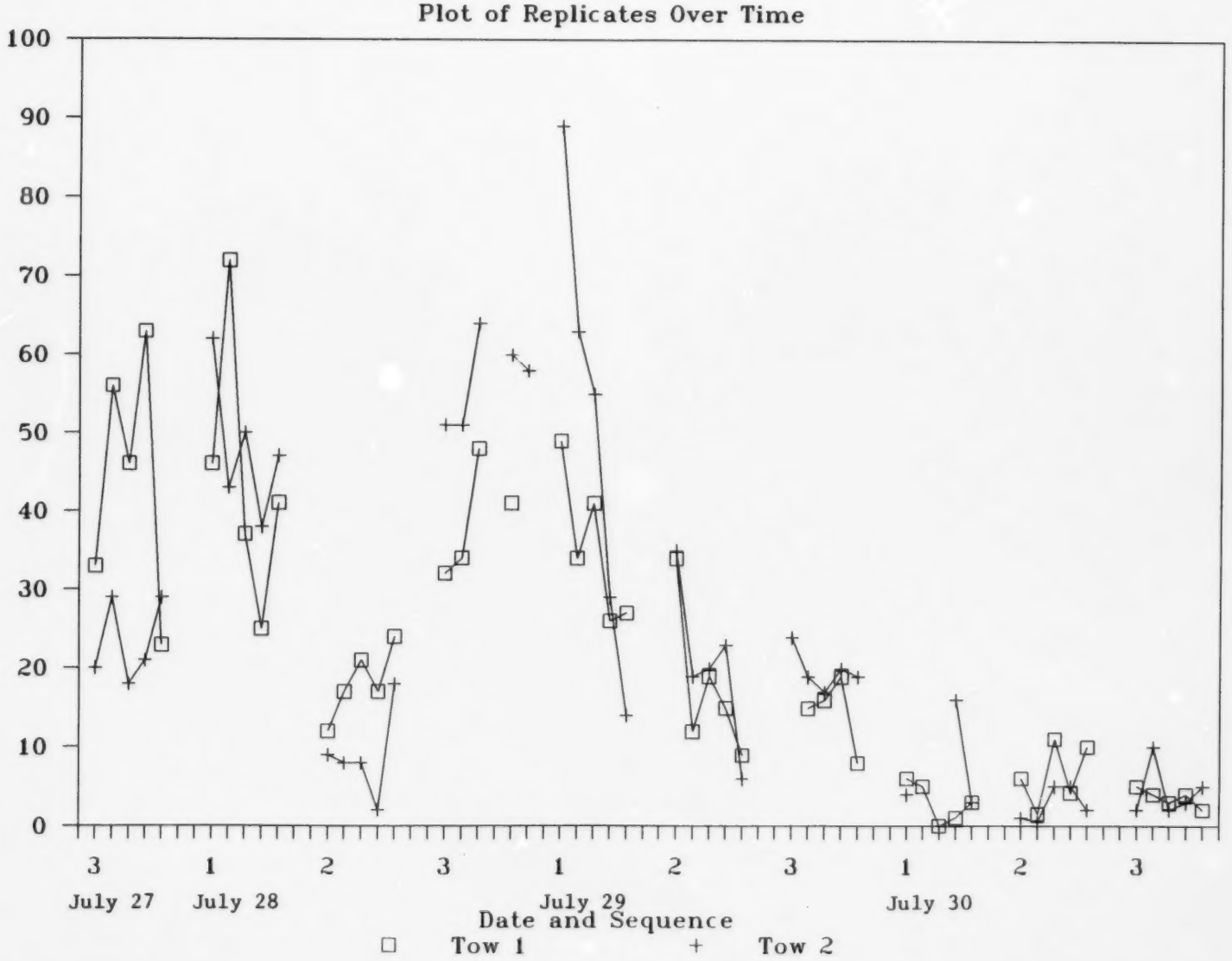
Plot of Replicates Over Time
Number of Larvae Per Tow
100
90
80
70
60
50
40
30
20
10
0
3
1
2
3
1
2
3
1
2
3
July 27
July 28
July 29
July 30
Date and Sequence
Tow 1
Tow 2

TABLE 1
Counts of Fish Larvae for Ten Minute Tows
From Paired 0.5 Meter Larval Tow Nets

Date	Time	Count Tow1	Count Tow2
July 27	14:54	33	20
	15:04	56	29
	15:16	46	18
	15:27	63	21
	15:38	23	29
July 28	09:20	46	62
	09:31	72	43
	09:43	37	50
	09:54	25	38
	10:05	41	47
	12:40	12	9
	13:10	17	8
	13:20	21	8
	13:35	17	2
	13:50	24	18
	15:00	32	51
	15:11	34	51
	15:22	48	64
	15:33	33 [1]	69 [2]
	15:44	41	60
	15:55		58
July 29	09:10	49	89
	09:21	34	63
	09:33	41	55
	09:45	26	29
	09:56	27	14
	12:02	34	35
	12:14	12	19
	12:25	19	20
	12:38	15	23
	12:50	9	6
	14:49	8 [3]	24
	15:00	15	19
	15:12	16	17
	15:25	19	20
	15:37	8	19
July 30	09:00	6	4
	09:11	5	5 [3]
	09:23	0	1 [3]
	09:35	1	16
	09:47	3	3
	12:01	6	1
	12:12 [4]	3	1
	12:34	11	5
	12:46 [5]	5	6
	12:59	10	2
	15:03	5	2
	15:14	4	10
	15:25	3	2
	15:36	4	3
	15:47	2	5

Notes: 1. Sample not properly preserved.
2. Bucket tipped; sample redone at 15:55 hours.
3. Live portion of sample not preserved; only couted dead larvae.
4. Twenty minute tow.
5. Twelve minute tow.

TABLE 2
Probability of Obtaining no Larvae in Sample For Various Sample Times Given the Expected Number of Larvae in a Ten Minute Sample

Expected no. of larvae in a ten minute sample	Sampling time in minutes				
	1	5	10	15	20
1	0.9048	0.6065	0.3679	0.2231	0.1353
2	0.8187	0.3679	0.1353	0.0498	0.0183
3	0.7408	0.2231	0.0498	0.0111	0.0025
4	0.6703	0.1353	0.0183	0.0025	0.0003
5	0.6065	0.0821	0.0067	0.0006	0.0000
6	0.5488	0.0498	0.0025	0.0001	0.0000
7	0.4966	0.0302	0.0009	0.0000	0.0000
8	0.4493	0.0183	0.0003	0.0000	0.0000
9	0.4066	0.0111	0.0001	0.0000	0.0000
10	0.3679	0.0067	0.0000	0.0000	0.0000
20	0.1353	0.0000	0.0000	0.0000	0.0000

TABLE 3
Means and Variances for Each Sequence and Tow Combination with Probabilities of Obtaining at Least One Zero Count

Date	Sequence	Tow	Mean of combination	Variance of combination	Prob.of at least one zero count
July 27	3	1	44.2	267.7	0.0000
	3	2	23.4	27.3	0.0000
July 28	1	1	44.2	301.7	0.0000
	1	2	48.0	81.5	0.0000
	2	1	18.2	20.7	0.0000
	2	2	9.0	33.0	0.0006
	3	1	38.8	52.9	0.0000
	3	2	56.8	32.7	0.0000
July 29	1	1	35.4	94.3	0.0000
	1	2	50.0	863.0	0.0000
	2	1	17.8	95.7	0.0000
	2	2	20.6	107.3	0.0000
	3	1	14.5	21.7	0.0000
	3	2	19.8	6.7	0.0000
July 30	1	1	3.0	6.5	0.2254
	1	2	7.7	52.3	0.0023
	2	1	6.5	15.8	0.0075
	2	2	2.7	4.7	0.2938
	3	1	3.6	1.3	0.1294
	3	2	4.3	11.3	0.0693

TABLE 4
Analysis of Variance to Determine
Differences Among Tows, Days and Sequences

SAS

General Linear Models Procedure

Dependent Variable: SCOUNT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	326.0976542	36.2330727	34.42	0.0001
Error	76	80.0111098	1.0527778		
Corrected Total	85	406.1087640			

R-Square	C.V.	Root MSE	SCOUNT Mean
0.802981	24.32537	1.026050	4.21802326

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TOW	1	1.1113755	1.1113755	1.06	0.3075
DAY	2	222.7220702	111.3610351	105.78	0.0001
SEQ	2	44.4999492	22.2499746	21.13	0.0001
DAY*SEQ	4	53.2010301	13.3002575	12.63	0.0001

TABLE 5
Analysis of Variance
Omitting Tow as a Factor

SAS

General Linear Models Procedure

Dependent Variable: SCOUNT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	324.9862787	40.6232848	38.56	0.0001
Error	77	81.1224853	1.0535388		
Corrected Total	85	406.1087640			

R-Square	C.V.	Root MSE	SCOUNT Mean
0.800244	24.33416	1.026420	4.21802326

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DAY	2	225.1408193	112.5704097	106.85	0.0001
SEQ	2	44.1729027	22.0864514	20.96	0.0001
DAY*SEQ	4	53.7246008	13.4311502	12.75	0.0001

TABLE 6
Individual Analysis of Variance for Each Day to Determine Differences Among Sequences

SAS

General Linear Models Procedure

---------- DAY=2 ----------

Dependent Variable: SCOUNT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	70.67834215	35.33917107	38.56	0.0001
Error	26	23.83078889	0.91656880		
Corrected Total	28	94.50913103			

---------- DAY=3 ----------

Dependent Variable: SCOUNT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	30.21768732	15.10884366	10.24	0.0005
Error	26	38.34529889	1.47481919		
Corrected Total	28	68.56298621			

---------- DAY=4 ----------

Dependent Variable: SCOUNT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.03354536	0.01677268	0.02	0.9781
Error	25	18.94639750	0.75785590		
Corrected Total	27	18.97994286			